ABSTRACT

The subject invention provides systems and methods that facilitate obfuscating a spam filtering system to hinder reverse engineering of the spam filters and/or to mitigate spammers from finding a message that consistently gets through the spam filters almost every time. The system includes a randomization component that randomizes a message score before the message is classified as spam or non-spam so as to obscure the functionality of the spam filter. Randomizing the message score can be accomplished in part by adding a random number or pseudo-random number to the message score before it is classified as spam or non-spam. The number added thereto can vary depending on at least one of several types of input such as time, user, message content, hash of message content, and hash of particularly important features of the message, for example. Alternatively, multiple spam filters can be deployed rather than a single best spam filter.